In-House Employee Policy Chatbot Using RAG

Task Description

The task was to design and implement a basic in-house chatbot capable of answering employee questions based on an internal employee policy document.
The chatbot uses a Retrieval-Augmented Generation (RAG) approach to ensure that responses are grounded in official policy content rather than generated from general model knowledge.
The system was implemented using the LangChain framework and a large language model, with a focus on clarity, correctness, and reproducibility.

Objectives

- Load and process an internal employee policy document
- Convert the document into searchable vector embeddings
- Retrieve relevant policy sections in response to user queries
- Generate accurate, policy-based answers
- Prevent hallucinated or unsupported responses

Approach

1. The employee policy document was loaded and split into manageable text segments.
2. Each segment was converted into a vector embedding and stored in a vector database.
3. When a user submits a question, the system retrieves the most semantically relevant policy segments.
4. Retrieved content is passed to a language model as contextual input.
5. The model generates a response strictly based on the retrieved policy information.

Deliverables

- A functional RAG-based chatbot capable of answering questions about the employee policy
- A document ingestion and preprocessing pipeline
- A vector database supporting semantic retrieval
- Prompt-controlled responses grounded in source material
- Source code demonstrating the full RAG workflow
- Documentation explaining the system design and usage

Skills Demonstrated

- Retrieval-Augmented Generation (RAG) architecture
- Practical use of LangChain for AI application development
- Semantic search using vector embeddings
- Prompt engineering for controlled and grounded responses
- Secure handling of API credentials
- Debugging and integration of third-party AI services

Outcome

The final system successfully answers employee queries using retrieved policy content, demonstrating an effective application of Retrieval-Augmented Generation for internal knowledge management.